

El Proyecto Qualitas: integración y análisis de procesos estadísticos

Cristina Rovira, Josep Ma. Martínez, Carme Saborit y Frederic Udina

Instituto de Estadística de Cataluña (Idescat)

Resumen

Para conseguir la máxima eficiencia en la producción estadística, es necesario que toda la información disponible trabaje conjuntamente en la descripción de una realidad social y económica compleja y cambiante.

Idescat viene trabajando en la modernización de la producción estadística oficial en dos líneas: la integración de las diversas fuentes de información estadística hacia un sistema integrado y en la documentación de los procesos estadísticos adaptados a este entorno de información integrada, que se ha denominado Proyecto Qualitas.

En una primera fase, el Proyecto Qualitas ha efectuado una descripción completa y detallada de los procesos en los que está trabajando el Idescat, tomando como guía el modelo GSBPM, un estándar internacional desarrollado y adoptado por la OCDE, el Eurostat y la UNECE. A partir de setenta proyectos ya documentados, se está trabajando en su transformación e integración en la plataforma tecnológica del Idescat (Plataforma Cerdà). El objetivo del Proyecto Qualitas es doble: por un lado, utilizar la documentación detallada de los procesos para garantizar la calidad y, por otro lado, proporcionar una guía y un control, a diferentes niveles, de aspectos como la automatización de las fases del proceso que lo permitan, las personas involucradas y la mejora y aprovechamiento de sinergias entre los diferentes procesos.

Palabras clave: GSBPM, procesos estadísticos, modernización de la producción y difusión estadística.

Sumario

1. Introducción	2
2. Sistema integrado de información estadística	3
3. El Proyecto Qualitas	5
4. Próximos pasos	13
5. Conclusiones	14
6. Referencias	15

1. Introducción

El propósito de las estadísticas oficiales es proporcionar estadísticas relevantes de alta calidad para la sociedad que permita la mejora de las políticas públicas y la toma de decisiones. Un enfoque moderno requiere la integración de información de muchas fuentes, alta calidad de los procesos de producción y el uso de múltiples canales de difusión para llegar a una amplia variedad de usuarios, lo que requiere un control constante y estricto de la seguridad y confidencialidad de los datos.

Durante los últimos tres años, el Idescat ha llevado a cabo un proceso de innovación en dos vías para alcanzar estos objetivos: tecnológica y de producción. Por un lado, en la vía tecnológica, la atención se ha centrado en la implementación de un nuevo sistema de gestión de la información (Plataforma Cerdà) basado en cuatro etapas conceptuales: captura de datos, integración de datos, generación de productos estadísticos y sistema de difusión. Esta plataforma tecnológica va a permitir la integración de los datos de entrada (ya validados en una primera etapa) por medio de la codificación de la información sobre las personas, entidades económicas y territorio a partir de la elaboración de tres registros estadísticos básicos. Cada pieza de información integrada estará disponible para los estadísticos a diferentes niveles, previo uso de un proceso de anonimización.

Por otro lado, y más recientemente, en la vía de los procesos de producción, la atención se ha centrado en la implementación de un sistema de metadatos de documentación de la producción basado en el GSBPM (*Generic Statistical Business Process Model*), que se ha denominado Proyecto Qualitas y que orienta la modernización de los procesos estadísticos;

Desde una perspectiva de negocio estadístico, el Proyecto Qualitas ofrece un estándar metodológico que va a permitir el control de la calidad y de la eficiencia de los procesos estadísticos y la posibilidad de compartir metadatos y de detectar mejoras y puntos críticos. Este Proyecto constituye una herramienta eficaz para la toma de decisiones y la determinación de las medidas necesarias para que los flujos de información y los procedimientos aplicables a los ficheros de datos que se utilizan para hacer estadísticas queden reflejados en documentos y protocolos que sean conocidos, aceptados y aplicados por todas las unidades del Instituto interesadas en cada caso, de acuerdo con las buenas prácticas estadísticas aplicables y respetando siempre los principios éticos y legales de privacidad y confidencialidad.

2. El Sistema Integrado de Información Estadística

Un aspecto importante de la modernización de la producción estadística es la integración de información que permita la reutilización eficiente de la misma y la combinación de diferentes fuentes de datos. El modelo basado en procesos específicos (*stove pipe*), en el que cada operación estadística se realiza sin ninguna conexión con cualquier otra (la figura 1 ofrece una visualización de este modelo) necesita ser reemplazado por otro en el que muchas fuentes de datos pueden ayudar a construir modelos estadísticos más complejos y generar nueva información de manera más eficiente (figura 2).

Figura 1.- Gestión de la información: esquema previo al proyecto

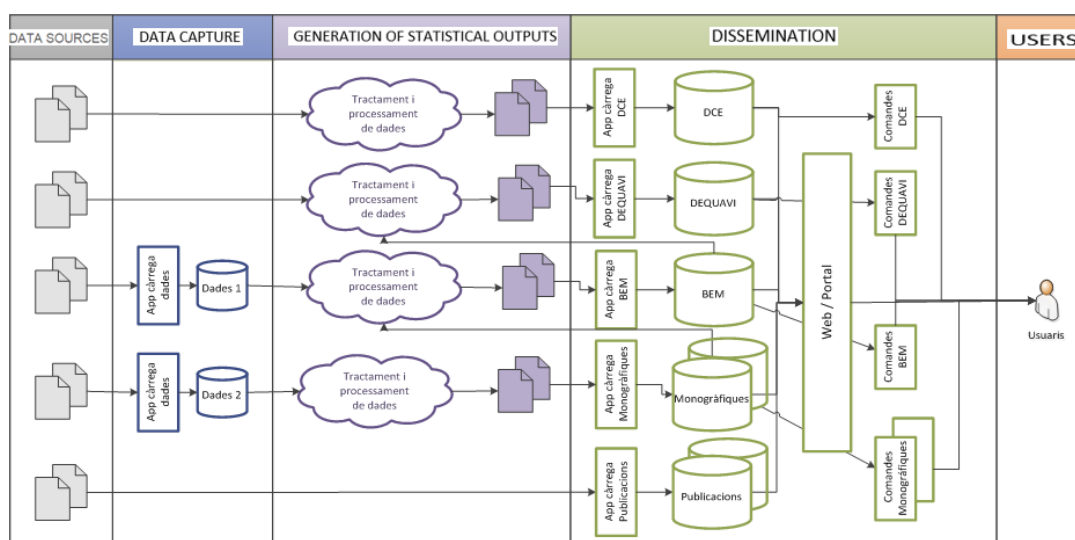
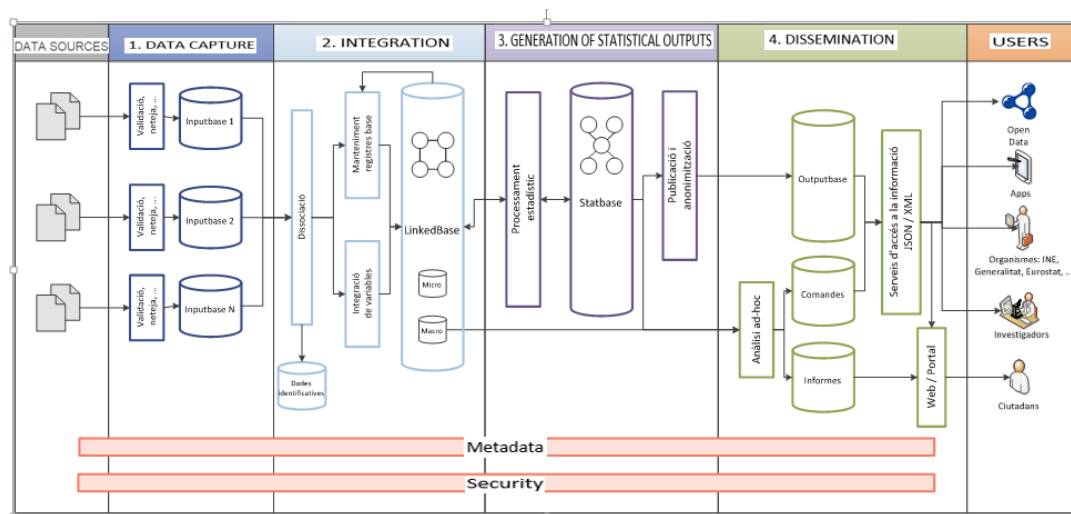


Figura 2.- El nuevo esquema de gestión de la información: la plataforma Cerdà.

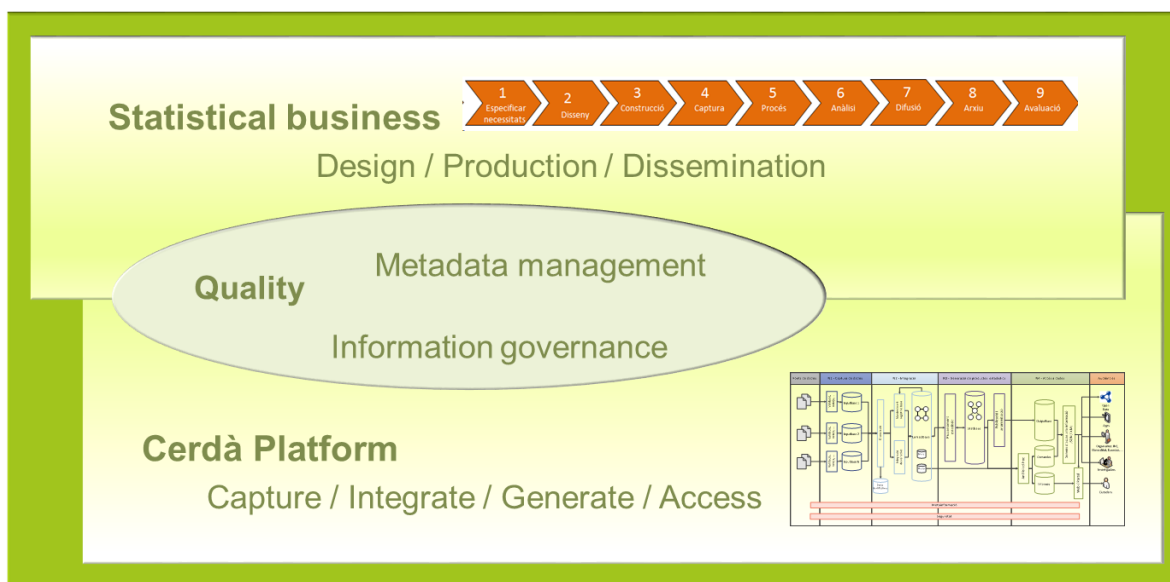


Nuestro enfoque de integración de la información se basa en Wallgren y Wallgren, 2014, aunque sin pretender la plena aplicación de su modelo. El núcleo del sistema se focaliza principalmente en la construcción de tres registros estadísticos básicos: población, entidades y territorio. Estos registros van a constituir la base del sistema integrado de información estadística elaborado principalmente a partir de información administrativa. En estos registros se reemplaza toda la información sobre las personas, las entidades (empresas, establecimientos...) y el territorio con información codificada. Cada individuo tendrá un código para permitir la vinculación de la información y evitar trabajar con campos identificables. Cualquier dato de identificación se mantiene separado y protegido de forma segura mediante el acceso restringido y controlado.

A partir de la codificación de los campos principales, la vinculación de la información debe ser fácil cuando una determinada operación estadística requiere el uso de datos de diversas fuentes. De esta manera, se facilita la integración y la reutilización de datos principalmente administrativos, y cuando sea necesario, por ejemplo, también datos de encuestas o de *big data*.

Para alcanzar la calidad y la eficiencia en un sistema integrado de información estadística el Idescat ha puesto en marcha dos planes directores estratégicos tal y como se ha comentado en la introducción y que se ilustran en la figura 3.

Figura 3. El Sistema Integrado de Información Estadística



3. El Proyecto Qualitas

El nuevo modelo integrado de producción estadística hacia el que avanza el Idescat, junto con una mayor presión en la producción y en un entorno de recursos económicos y humanos más reducidos, hace que resulte necesario la sistematización del conocimiento de los procesos estadísticos. En este contexto de cambio hacia un modelo más eficiente nace el Proyecto Qualitas con la misión de conocer y documentar de manera estandarizada los procesos estadísticos en el Idescat.

El objetivo del Proyecto Qualitas es doble: por un lado, utilizar la documentación detallada de los procesos para garantizar la calidad y, por otro lado, proporcionar una guía y un control, a diferentes niveles, de aspectos como la automatización de las fases del proceso que lo permitan, las personas involucradas y la mejora y aprovechamiento de sinergias entre los diferentes procesos.

Este proyecto ha constado de dos fases: en la primera, se recoge una amplia y estandarizada documentación relativa a los procesos estadísticos actualmente vigentes en el Instituto: cómo llegan los datos, quién interviene, cómo se procesan, qué aplicaciones se utilizan, cómo se difunden, a través de qué productos de difusión, etc. En la segunda fase, los procesos anteriores deben ser analizados con el fin de mejorar la eficiencia y proceder a su integración en el nuevo paradigma de la producción.

En la primera fase del Proyecto Qualitas se han considerado los siguientes aspectos:

1. Identificación de los procesos estadísticos
2. Programación y realización de reuniones de trabajo con las personas clave
3. Recogida y análisis de la información
4. Documentación del proceso estadístico
5. Validación del proceso y obtención de la versión definitiva

El punto de partida fue el análisis de las actuaciones estadísticas responsabilidad del Idescat recogidas en el Programa anual de actuación estadística de Cataluña del 2015¹, año de referencia del Proyecto Qualitas. Se identificaron inicialmente un conjunto de 108 actuaciones estadísticas

¹ Decreto 165/2014, de 23 de diciembre, por el que se aprueba el Programa anual de actuación estadística para el año 2015

susceptibles de ser documentadas de un total de 124, de las cuales se priorizaron 88, que a su vez se agruparon en 73 procesos estadísticos.

Se han llevado a cabo un total de 108 reuniones de trabajo en las que han participado un total de 48 personas del Idescat (principalmente técnicos de las áreas de producción). Las reuniones han constituido un verdadero centro de conocimiento y han permitido aunar y compartir toda la información necesaria para el desarrollo de los documentos estadísticos, desde la recopilación de la información hasta su difusión, dando valor añadido al análisis del proceso y permitiendo así poner en común diferentes perspectivas de trabajo. Se han priorizado y programado simultáneamente operaciones de naturaleza análoga e identificado las personas claves en el desarrollo del proceso estadístico cuya implicación ha sido fundamental para el éxito del proyecto.

La información recogida para la documentación de cada proceso estadístico se resume en la figura 4.

Figura 4. Un ejemplo de tabla de contenidos del proceso de documentación

Tabla de contenidos	
1	Tablas resumen del proceso
1.1	Datos identificativos
1.2	Datos genéricos
2	Tablas resumen del subproceso
2.1	Datos identificativos
2.2	Datos genéricos
3	Diagrama de flujo de datos
4	Desarrollo del proceso
5	Modelo GSBPM
6	Puntos críticos del proceso
7	Definiciones y abreviaturas
8	Documentos relacionados
9	Documentos resultantes
10	Modelos de documento
11	Anexo

El primer punto, *tablas resumen*, recoge información sobre las personas responsables del proceso y de cada una de sus fases, que se registra en la sección *datos identificativos*. Por su parte, en la sección de *datos genéricos* se recogen las leyes y otras normas aplicables, así como una descripción resumida del proceso y sus fases. En algunos casos, se identifica y describe un

subproceso. El *flujo de datos* e información para apoyar el proceso se han documentado, también, cuidadosamente.

En la sección 4 *desarrollo del proceso* se proporciona una descripción más detallada de cada paso del proceso. La sección 5 está dedicada a la correspondencia de las diferentes fases con el modelo GSBPM adoptado por la UNECE, Eurostat y otras oficinas de estadística de referencia. En nuestra documentación relacionamos las fases y pasos que se describen con los considerados como estándar por el GSBPM.

Una cuestión clave en la calidad de los procesos estadísticos (véase Ehling et al, 2007) es la identificación de los *puntos críticos del proceso* y la sección 6 está dedicada a ello. El resto de la tabla de contenidos incluye referencias a otra información de utilidad.

Se detalla a continuación la información más relevante recogida en la tabla de contenidos de la figura 4.

Tablas resumen del proceso

Se recogen los datos más significativos del proceso: datos identificativos y datos genéricos (datos de la fase de recogida, de la fase de proceso y de la fase de difusión).

Datos identificativos.

Incluyen la identificación del proceso, su correspondencia con la actuación estadística oficial, en qué unidad y técnico recae la responsabilidad del proceso, una breve descripción y las variables principales a investigar.

Datos identificativos	
Código - Nombre	
Código PAAE	
Operación estadística	
Subdirección responsable del proceso	
Responsable subdirección	
Área responsable del proceso	

Responsable de área	
Responsable técnico proceso	
Descripción	
Variables principales a investigar	
Observaciones	

Datos genéricos.

Incluyen la normativa aplicable, el origen de los datos, información relativa a las fases de recogida, proceso y difusión, así como la dependencia con otros procesos.

Datos genéricos		
Normativa genérica aplicable		
Origen de la información	Tipo de datos de origen	
	Fuente concreta	
	Organismo asociado	

FASE RECOGIDA (Grupo múltiple)				
Área responsable				
PRIMER FICHERO –				
Datos iniciales				
Tipo de entrada				
Nombre de fichero		¿Encriptado?		
Periodicidad		¿Previa solicitud?		
Canal entrada datos				
Ubicación datos				
Formato datos		¿Microdatos?	¿Anonimizado?	-
Persona Organismo contacto	/ de -			

Datos proceso					
Ubicación					
Software		Lenguaje			
Observaciones					
Datos finales					
Ubicación					
¿Esquema de BD?		Nombre esquema		Aplicación web	
Formato datos		¿Microdatos?			Anonimizado?

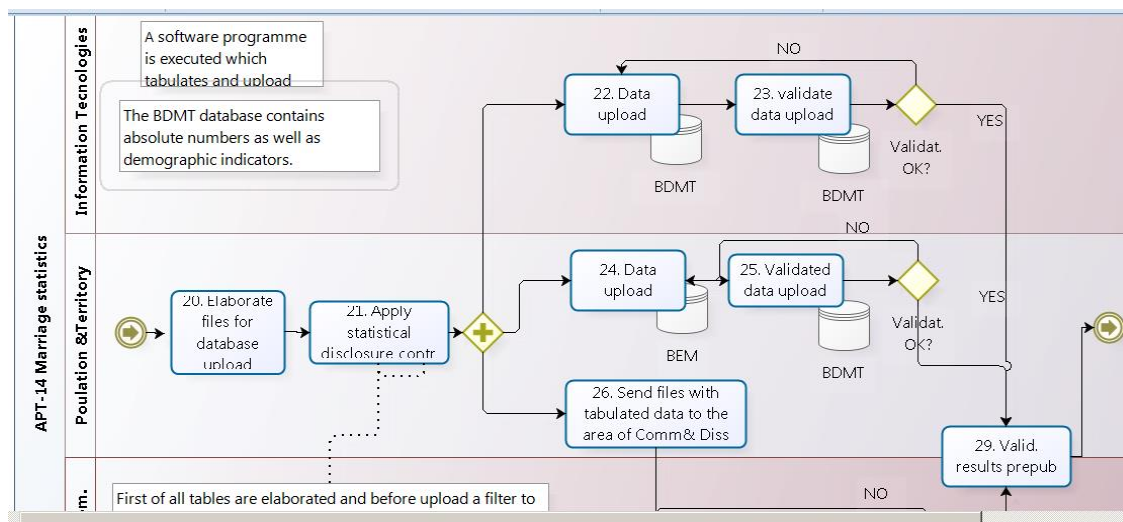
FASE PROCESO (Grupo múltiple)					
Área responsable					
FICHEROS					
Datos proceso					
Nombre de ficheros					
Ubicación					
Software		Lenguaje			
¿Uso de estadística instrumental?			¿Aplica secreto estadístico?		
Secreto estadístico					
Significación estadística					
Observaciones					
Datos resultado final -					
Ubicación					
Esquema de BD?		Nombre esquema		Aplicación web	
Formato datos		¿Microdatos?			¿Anonimizado?
Datos resultados intermedios (datos no directamente publicables)					
Nombre fichero					
Ubicación					
Formato datos		¿Microdatos?			¿Anonimizado?

FASE DIFUSIÓN					
Área responsable					
PREPARACIÓN de los datos					
Datos proceso					
Ubicación					
Software		Lenguaje			
Datos resultado final					
Ubicación					
Esquema de BD?		Nombre esquema		Aplicación web	
Formato datos		¿Microdatos?			¿Anonimizados?
DIFUSIÓN de los datos					
Aplicaciones web					
Productos de comunicación					
Otros datos					
Dependencias con otras estadísticas oficiales	Productos estadísticos oficiales				
	Otros				

Diagrama de flujo de datos.

Muestra, mediante un esquema gráfico, la secuencia de pasos del proceso estadístico a realizar para obtener el resultado de la operación estadística, desde su inicio hasta su final, así como la unidad de la organización responsable de cada paso.

Figura 5. Ejemplo de flujo de datos, especificando las unidades a cargo de cada paso



Desarrollo del proceso.

Se describe detalladamente cada una de las tareas o actividades incluidas en el diagrama de flujo anterior y se asignan áreas responsables.

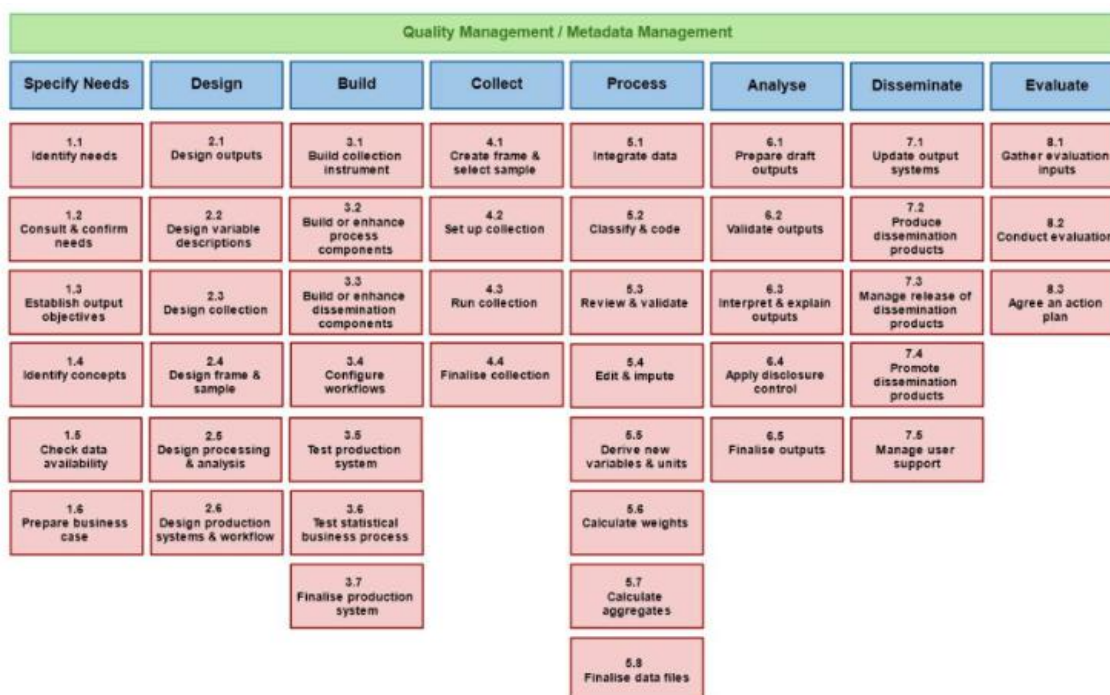
Modelo GSBPM.

El objetivo del modelo GSBPM es recoger información de todas las fases que intervienen en el proceso productivo de manera estandarizada, favoreciendo, así, su sistematización. Es un modelo de referencia flexible, estructurado en tres grandes niveles, que se ha adaptado a los requerimientos y contexto específico del Idescat:

- Nivel 0: se identifica el proceso estadístico
- Nivel 1: incluye 8 fases, distribuidas en 3 ámbitos:
 - Planificación (necesidades, diseño, construcción)
 - Ejecución (recogida, proceso, análisis, difusión)
 - Evaluación (evaluación)
- Nivel 2: desglosa en 44 subprocesos las 8 fases del nivel 1.

En la figura 6 se pueden consultar las 8 fases y los 44 subprocesos del modelo GSBPM.

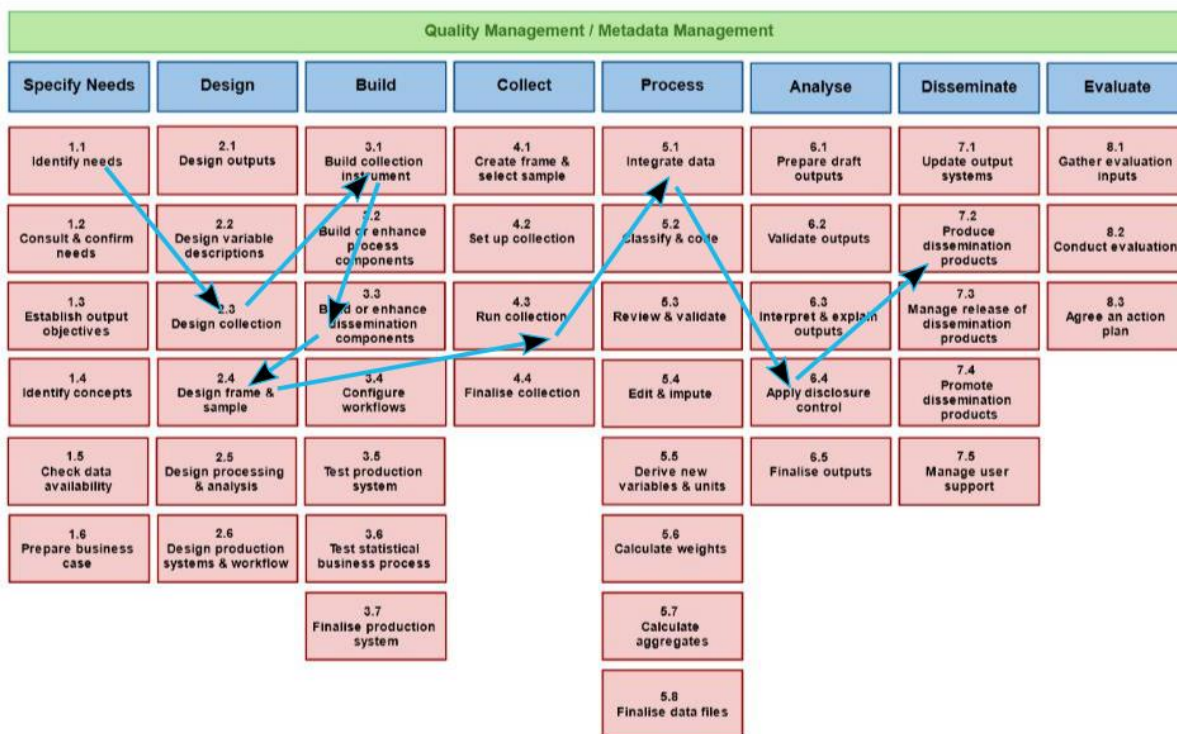
Figura 6. Niveles 1 y 2 del modelo GSBPM 5.0.



El modelo GSBPM tiene la peculiaridad de que se adapta a las características de cada proceso de producción estadística. La secuencia de un proceso no es lineal, es decir, los subprocesos

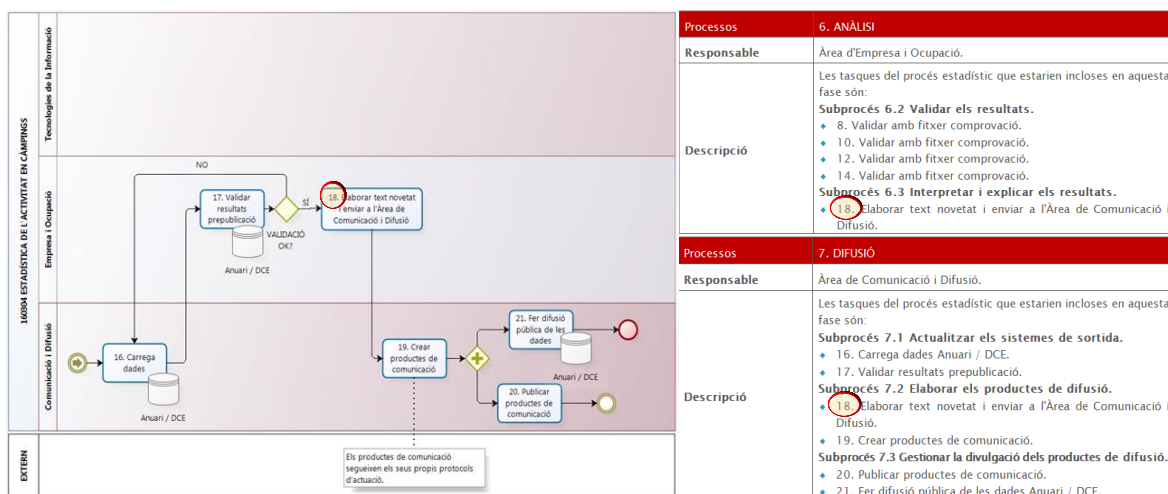
no tienen por qué seguir el mismo orden ni secuencia, ni tienen que estar siempre presentes todos los subprocesos.

Figura 7. Ejemplo de secuencia de un proceso estadístico.



Las actividades o tareas identificadas en el diagrama de flujo se han asignado a los subprocesos descritos en el nivel 2 (en la figura 8 se puede visualizar un ejemplo).

Figura 8. Ejemplo de secuencia de un proceso estadístico.



4. Próximos pasos

- Asegurar que los procesos están siempre vigentes. En este sentido hay que establecer un protocolo para la documentación del Proyecto Qualitas y también un sistema de información para centralizar y explotar la información de la ficha técnica de los procesos, así como para controlar el cambio de versiones. En este sentido, se está trabajando en la construcción de una herramienta de gestión de documentos para mantener toda la información actualizada y que a su vez permita una óptima gestión de la misma. Este sistema de metadatos ha de evolucionar para convertirse en la columna vertebral común por lo que respecta a la Plataforma Cerdà.
- Elaboración de un nuevo protocolo de actuación para la planificación, el desarrollo y la producción estadística del Idescat y análisis del impacto de la migración de estadísticas a la Plataforma Cerdà y de la implantación de nuevas estadísticas en el nuevo modelo de producción.

Recientemente se ha procedido a un doble ejercicio sobre la Plataforma Cerdà: como ejercicio inicial, se ha llevado a cabo una prueba de concepto para la elaboración de una estadística consolidada en la nueva Plataforma; en un segundo estadio, se han desarrollado nuevas estadísticas directamente en la Plataforma. Ambos ejercicios, van a ser muy útiles para evaluar cómo incide el nuevo modelo de producción en el flujo de actividades del Instituto, qué impacto tiene en los roles e intervenciones de las diferentes unidades y si existen modificaciones en lo que respecta a los puntos críticos del proceso. El estándar GSBPM y los resultados de esos análisis deben guiar la definición pautada de un nuevo protocolo de actuación para el desarrollo y elaboración de estadísticas del Idescat, ayudando, además, a identificar procesos o tareas comunes en vistas a su homogeneización, estandarización y reutilización.

- La segunda fase del Proyecto Qualitas consiste en estudiar, analizar y convertir, uno por uno, los procesos estadísticos al nuevo paradigma por el cual, el resultado estadístico se produce utilizando toda la información disponible relacionada, partiendo de la base de los registros estadísticos, y permitiendo que los resultados de cada proceso estén disponibles para productos estadísticos posteriores. Esto nos permitirá lograr aspectos relevantes de acuerdo con el código de buenas prácticas de las estadísticas europeas: metodología sólida, procedimientos estadísticos adecuados, carga no excesiva para los encuestados y relación coste eficacia.

5. Conclusiones

Los trabajos desarrollados en el Proyecto Qualitas han permitido dotar a la organización de un sistema estructurado y estandarizado de documentación de los procesos estadísticos. El uso del modelo GSBPM supone un avance para la implementación de los nuevos modelos de producción estadística basados en la integración de datos. Por un lado, la sistematización confiere garantías de continuidad a los procesos y a su reproducción, independientemente de cambios ajenos a la naturaleza de los mismos (por ejemplo, la movilidad de los técnicos). Por otro lado, la estandarización permite la comparabilidad de procesos y una mejor transmisión del conocimiento.

Se ha asegurado la gestión del conocimiento mediante la homogeneización y sistematización de la documentación de los procesos estadísticos, estandarizando lenguaje, acciones, etc. Un total de 73 procesos estadísticos han sido documentados de manera estandarizada y homogenizada a partir de una ficha técnica común que va a posibilitar el análisis y la evaluación de los procesos estadísticos. Tener los procesos documentados es el punto de inicio para un sistema más eficiente y de mejora continua. Algunas utilidades se concretan en:

- la detección de duplicidades y posibles simplificaciones en los procesos.
- disponer de una lista actualizada de las herramientas usadas
- la identificación de los archivos y programas importantes,
- aumentar las garantías de continuidad de un proceso estadístico
- posibilitar la comparación de diversos procesos estadísticos
- mejorar la toma de decisiones sobre un proceso estadístico
- distribuir de manera más eficiente los recursos
- identificar procesos que no estaban suficientemente documentados
- minimizar las inconcreciones
- mejorar las buenas prácticas

Desde el punto de vista de implicación del personal del Idescat, las tareas llevadas a cabo en el desarrollo del proyecto han fomentado la participación, la transmisión de conocimiento, la implantación de modos sistemáticos de trabajar y la generación de sentimiento de pertenencia.

6. Referencias

Ehling, M. and Körner, T. (eds.) (2007). Handbook on Data Quality Assessment Methods and Tools. Eurostat.

European Statistical System (2011) European Statistics Code of Practice - revised edition 2011

UNECE (2013). The Generic Business Process Model v 5.0.

<http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0> (as found in April 2016)

Wallgren A. and Wallgren B. (2014), Register-based statistics. Statistical Method for Administrative Data, 2nd edition. J. Wiley.